The Information Manifold for Relatively Bounded Potentials

R. F. Streater
Dept. of Mathematics, King's College London,
Strand,London WC2R 2LS

2/8/1998

Abstract

We construct a Banach manifold of states, which are Gibbs states for potentials that are form-bounded relative to the free Hamiltonian. We construct the (+1)-affine structure and the $(+1)-$connection.
keywords Information manifold, Fisher metric, quantum geometry, Bogoliubov metric, Kubo theory, statistical manifold.

1 Introduction

The ultimate goal of the present project is a quantum version of the theory of information (or statistical) manifolds; in classical probability this circle of ideas has been rather successful in many fields from estimation theory to dissipative dynamics in neural networks [2]. We were inspired by the nice work of Pistone and Sempi, who put the classical theory on a firm mathematical foundation in the infinite-dimensional case [33]. While our ambitions are the same as [33], our results are not as complete, and the technical problems, arising from the non-commutative nature of the potentials, are quite different. The problem first arose in the quantum theory of many-body systems at 'finite' temperature, in the works of Matsubara, Mori and Kubo [25, 26, 23]. There, we find the correlation functions for observables written as imaginary-time-ordered products of operators. The mathematical theory was advanced by Bogoliubov, [6] who showed that the two-point correlation function was real and positive-definite. In the geometrical context, this is thus a Riemannian metric on the vector space of Hermitian operators; it is often known as the Bogoliubov-Kubo-Mori, or BKM, metric.

Parametric estimation theory starts with a family of probability distributions, and also the data which give us a distribution in the form of a histogram. We seek the best representation the data; this is done by finding the member of the family that minimises the 'distance' to the distribution of the data; here, which concept of distance to be used is one of the problems. Gauss used the Euclidean distance, giving his famous least-squares fit to the data. This distance, however, depends on the coordinates used. Fisher [11] introduced the *information matrix*, which is a tensor under change of coordinates, and this was developed into a Riemannian metric tensor on the manifold of parametrised distributions by Rao [34]. Dawid [10] realised that the theory also needed an affine connection, and that this did not have to be that of Levi-Civita. Ideas from information theory were then incorporated, and the 'distance' to be minimised turned out to be the Kullback-Leibler relative entropy. The poetic geometry involving the dual affine structures was then put together by Amari, in a notable book [1]. The manifold of states is determined by a chosen subspace $\mathcal{X}$, spanned by (linearly independent) random variables $\{X_1, \ldots, X_n\}$. It can be parametrised by the canonical coordinates ξ^j, $\xi \in V$, where V is a convex cone in $\mathbf{R}^n$. Let $X = \sum \xi^j X_j$; the *exponential family* determined by $\mathcal{X}$ is the set of states of the form

$$p_{_X} = Z^{-1} \exp\{-\sum \xi^j X_j\} = \exp -\{X + \Psi(\xi)\}. \tag{1}$$

The canonical coordinates are (inverses of) generalised temperatures [19] and the Massieu function $\Psi = \log Z$ is the thermodynamic potential. The $(+1)$-affine structure comes from forming convex mixtures of the ξ^j; that is, the mixtures of $p_{_X}$ and $p_{_Y}$ are states of the form $p_{_{\lambda X+\lambda' Y}}$. In this paper, λ' will denote $1 - \lambda \geq 0$. Since the ξ are global affine coordinates, this affine structure is flat and torsion-free. The Legendre dual to Ψ is the entropy, written as a function of the probability, and it also plays the role of a thermodynamic potential, The dual coordinates,

$$\eta_i = -\frac{\partial \Psi}{\partial \xi^i} \tag{2}$$

are the expectation values of the random variables X_i in the state $p_{_X}$.

The η_i are also global coordinates, and the (-1)-affine structure is defined by forming convex sums of these coordinates; it is therefore also flat and torsion-free. It coincides with the usual convex mixture of states, and so is called the *mixture* affine structure. Each of these affine structures defines a concept of parallel transport of vectors in the tangent space; neither

of these affine structures is metric invariant, but they are dually related by the metric. Amari also interpolates between these, to get the family of (α)-affine structures, of which $\alpha = 0$ is self-dual and therefore metric: it is the Levi-Civita affine structure.

It has been remarked [20, 22, 18, 4, 2, 39] that several dissipative models used in neural nets and physics can be expressed as the projection or rolling of a linear dynamics onto the surface given by a family of distributions. The random variables whose distributions form the manifold are taken to be the slow variables of the theory; the other variables are thermalised by the projections that keep the dynamics on the manifold.

The mathematical legitimacy of the procedure was strengthened by the studies of Chentsov [8]. He had the idea of regarding stochastic maps as the natural morphisms between statistical structures. The Fisher-Rao metric is contracted by any such map; moreover, it is the only Riemannian metric (up to a constant factor) to have this property. In physics, dissipative dynamics is given by a semigroup of stochastic maps, and the contractive property is the expression of convergence to equilibrium at large times. These are necessary properties of any good theory. Thus there is a certain uniqueness, in the classical case, of the geometry of parametric families. Chentsov remarked that this is not the case in the geometry of quantum information manifolds in n dimensions. This was studied by Hasagawa, Nagaoka and Petz [14, 15, 32, 16, 17, 27, 29]. See also [28]. The set of faithful density matrices is a manifold $\mathcal{M}$ of dimension $n^2 - 1$. The morphisms are taken to be stochastic completely positive maps. Chentsov showed that there are many metrics on the tangent space $\mathcal{T}$ of $\mathcal{M}$ that are contracted by these morphisms. One can identify $\mathcal{T}$ at ρ with the linear space of Hermitian matrices with zero expectation in the state ρ. This is the quantum analogue of the 'centred random variables' that make up the tangent space in $+1$ coordinates in the classical case. Chentsov's work on the possible metrics was completed by Petz [30] in the case of finite dimensions. Examples of these can be found in Roepstorff [36]. Hasagawa, and Nagaoka, in particular emphasise two important cases. These are the symmetric *KMS* metric, and the *BKM* metric.

Given a faithful density matrix, ρ, the *KMS* metric on the vector space of $n \times n$ Hermitian matrices is constructed from the complex scalar product $\langle X, Y\rangle = \mathrm{Tr}\,(\rho X^* Y)$ by taking the real part. The *KMS* metric has been used in quantum estimation theory [41, 7], and it coincides with the Study-Fubini metric on the projective sphere when restricted to the pure states. I have also used it extensively in [38]. In spite of this, it seems that the *BKM*

metric is better. It is defined on the space of *centred* operators $\widehat{X} = X - \rho.X$ by

$$\langle \widehat{Y}, \widehat{X} \rangle := \mathrm{Tr} \int_0^1 \rho^\lambda \widehat{Y} \rho^{\lambda'} \widehat{X} d\lambda \tag{3}$$

Here, and elsewhere in the paper, $\lambda' = 1 - \lambda$. First, in the quantum case, the $(\pm)$-affine structures are not dual relative to the *KMS* metric. Alternatively, one can take, as in [27], the mixture affine structure as a start; then its dual relative to the *KMS* metric is not flat and torsion free. It follows that there do not exist dual potentials, related by a Legendre transform, corresponding to the physically important objects, the entropy and the Massieu function. The second reason why I now prefer the *BKM* metric is mathematical; the *BKM* metric is smaller than the *KMS* metric; the latter does not exist in general for unbounded operators, and certainly not for forms. The mathematical *stylishness* of the *BKM* version of the information manifold is so compelling that perhaps the extensive work on quantum estimation should be redone with the *BKM* metric replacing the *KMS* metric.

In the classical case, Pistone and Sempi [33, 12] introduce information manifolds in general, not parametrised by a finite number of slow variables. Thus they are in the field of *nonparametric estimation theory*. However, they must start somewhere, and they fix a basic underlying measure space, whose measure μ need not be finite, but is used to specify the null sets. The probability measures of the manifold are then those smoothly related to the given one. The present paper is an attempt to get a nonparametric version in the quantum case. It follows up earlier work [37], in which the Hilbert space was of infinite dimension, but the manifold was of finite dimension, corresponding to limiting our attention to finitely many 'slow' variables. In the quantum case, we need a trace, not necessarily finite, to play the role of μ; we need a density matrix to play the role of p; this is provided by a 'free Hamiltonian' H_0, a positive selfadjoint operator with domain $\mathcal{D}(H_0)$, on a Hilbert space $\mathcal{H}$, such that there exist $\beta_0 > 0$ with

$$\rho_0 = Z_0^{-1} e^{-\beta H_0} \text{ is a density operator for all } \beta > \beta_0. \tag{4}$$

This condition holds for the harmonic oscillator, and also for the Laplacian in a compact region in $\mathbf{R}^n$, with smooth boundary and Dirichlet boundary conditions, or in a rectangle in $\mathbf{R}^n$ with periodic boundary conditions. In all these examples, $\beta_0 = 0$. The condition corresponds to a thermodynamically stable system in a finite box, in which there are no phase transitions for $\beta > \beta_0$. The zero-point energy of H_0 has no significance, as the addition

of a constant to H_0 is cancelled by the corresponding change in Z; we may therefore assume that $H_0 \geq I$. By scaling, we may assume that $\beta_0 < 1$, and we start with a state of the form

$$\rho_0 = e^{-(H_0+\Psi_0)} \tag{5}$$

Here, $\Psi_0 = \log Z_0$. The condition given by eq. (4) is enough to allow the quantum analogue of the 'Cramer class' of random variables u arising in the classical case [33].

In §(2) we shall construct our first patch of the manifold. It will be a set of states related to the basic state ρ_0 by a small form perturbation of H_0. A *form* is a bilinear real map $\varphi, \varphi \mapsto X(\varphi, \varphi) \in \mathbf{R}$, where φ runs over the form-domain $Q(X)$. For example, the positive selfadjoint operator H_0 defines the quadratic form q_0 with form domain $Q_0 = D(H_0^{1/2})$ by the definition

$$q_0(\varphi, \varphi) = \langle H_0^{1/2}\varphi, H_0^{1/2}\varphi \rangle. \tag{6}$$

The theory of small forms allows us to write the operator H_X as the unique selfadjoint operator satisfying

$$\langle H_X^{1/2}\varphi, H_X^{1/2}\varphi \rangle = q_0(\varphi, \varphi) + X(\varphi, \varphi), \tag{7}$$

for all $\varphi \in Q_0$. The perturbed state

$$\rho_X = Z_X^{-1} \exp -\{H_X\} = \exp -\{H_X + \Psi_X\} \tag{8}$$

is shown (lemma 4) to exist provided that the *form bound* of X is smaller than $1 - \beta_0$; it inherits most of the good properties of ρ_0. The forms X which are q_0-bounded give us the Cramer class at ρ_0.

States of the form ρ_X are the canonical states for the Hamiltonian $H_0 + X$. The case of bounded perturbations has been extensively analysed by Araki [3]. In linear response theory such states are thought of as the equilibrium state reached in response to an external field, whose 'effective potential' is X. We do not insist on this interpretation; for example, in the version of non-equilibrium statistical mechanics called 'statistical dynamics' [38] we regard ρ_X as a nonequilibrium state parametrised by X. Ingarden [18] has called the possible X the 'slow variables'; Jaynes calls them the accessible variables, in line with his subjective view of entropy. The point of working in infinite dimensions is to have a space of states large enough to contain the dynamics, so that the 'reduced description' can be given a

geometric flavour as the projection from the full manifold to a submanifold described by the means of the variables of interest.

The parametrisation of the perturbed states is established by an excursion into the theory of *sesqiforms*. We show that if X is relatively form-bounded, then the expectation value $\rho_0.X = \mathrm{Tr}\,(\rho_0 X)$ can be given an unambiguous meaning, and is continuous in X when the space of relatively bounded forms is provided with with a natural norm, in which it becomes a Banach space, $\mathcal{T}(0)$. A form obeying $\rho_0.X = 0$ is said to be *centred*. The subspace $\widehat{\mathcal{T}}(0) \subseteq \mathcal{T}(0)$ given by centred variables then defines a closed subspace; the open ball of radius $1-\beta_0$ in $\widehat{\mathcal{T}}(0)$ is in bijection with a collection of states of the form eq. (8). This is our first *patch* of the information manifold, i.e. mapping from the set of states into the open unit ball of a Banach space.

In §(3) we develop analysis with the first patch; in particular, we derive the Duhamel formula for the difference of two states of the form eq. (8) in terms of integrals of sesqiforms.

In §(4) we look at the two affine structures (± 1) in the first patch. Parallel transport in the $(+1)$ structure is easy to define; but the (-1)-affine sum, which is ordinary mixture of states, may lead outside the manifold.

In §(5) the manifold is extended by adding overlapping patches based on points in already established patches. The key here is that a perturbed state ρ_X inherits all the good properties of ρ_0, and that the norms on overlapping patches are equivalent. That we should be able to do better than just the first patch is easy to understand; if $X = (1/2)H_0$, then X is H_0-small, and we can define $H_X = H_0 + X = (3/2)H_0$. Then an operator Y can be H_X-small without $X+Y$ being H_0-small. So we define $H_0+X+Y = (H_0+X)+Y$ in two stages. In this case we get the same operator for $\beta(H_0+X+Y)$ as if we use $3/2\beta(H_0+2/3Y)$, so in this case we could get there in one step from a state of different temperature. However, in general we expect to reach new states, not obtainable in one step. In this way we build up our manifold of states, reachable from ρ_0 in a finite number of steps. It is clear that the whole space of H_0-bounded operators cannot be reached from the state of given beta; for $X = -H_0$ will never be reached; roughly, the manifold we construct lies in the direction of $+H_0$. The $(+1)$ affine structure and parallel transport can be extended to the whole manifold, which is proved to be convex.

2 Sesquiforms and Perturbation Theory

In this section we extend the definition of perturbed states, beyond that considered in [37], in two ways. First, we allow X to be a quadratic form, bounded relative to the quadratic form q_0. This means the following. Let X be a sesquilinear form defined on Q_0; it is said to be q_0-bounded if there exist numbers, a, b such that

$$|X(\psi,\psi)| \leq a q_0(\psi,\psi) + b\|\psi\|^2 \text{ for all } \psi \in Q_0. \tag{9}$$

If a can be chosen less than 1 (by a good choice of b), then we say that X is q_0-small. In our case, we shall need to choose a smaller than $1-\beta_0$. If a can be chosen to be arbitrarily small, we say that X is q_0-tiny.

It is not hard to show that $D(H_0) \subseteq Q_0$. Any H_0-small operator is also q_0-small [35], Th X.18.

The second extension of [37] is that we consider simultaneously the set of all q_0-bounded forms, and provide them with a norm; we obtain a parametrisation of the space of perturbed states by X of small norm by the open unit ball of a Banach space: this is our first patch of the manifold.

2.1 Sesquiforms

We can give a meaning to left and right products of quadratic forms by certain operators. Suppose that q is a quadratic form with domain $Q(q)$; then q defines a sesquilinear form $q(\phi,\psi)$ by polarisation, with domain $Q(q)\times Q(q)$. Let A, B be operators such that A^* and B are densely defined, A^* taking $D(A^*)$ into $Q(q)$, and B mapping $D(B)$ into $Q(q)$. Then by the expression AqB we mean the sesquilinear form given by

$$\phi,\psi \mapsto q(A^*\phi, B\psi), \qquad \phi \in D(A^*), \quad \psi \in D(B).$$

It is obvious that the product is associative: $(AB)q = A(Bq)$. More generally, given dense linear sets $\mathcal{D}_1$, $\mathcal{D}_2$, a sesquilinear map from $\mathcal{D}_1 \times \mathcal{D}_2$ to $\mathbf{C}$ will be called a *sesquiform*. A sesquiform is not required to be symmetric. The 'formal adjoint' of the sesquiform q is the form q^* with domain $D_2 \times D_1$, and given by

$$q^*(\varphi,\psi) = \overline{q(\psi,\varphi)}. \tag{10}$$

We note that the restriction of a sesquiform to a pair of dense linear subspaces of $\mathcal{D}_1$ and $\mathcal{D}_2$ is also a sesquiform, and that sesquiforms with the same domains can be added to give a sesquiform.

Definition 1 *A sesquiform q will be said to be bounded if there exists C such that*

$$|q(\varphi, \psi)| \le C\|\varphi\| \, \|\psi\| \text{ holds on the domain.}$$

Lemma 2 *Suppose that X is a q_0-bounded symmetric form defined on Q_0. Then $R_0^{1/2} X R_0^{1/2}$ is a bounded symmetric form defined everywhere. Conversely, if X is a symmetric form with domain Q_0, and $R_0^{1/2} X R_0^{1/2} < 1$, then X is q_0-small.*

PROOF.
Recall that we have normalised H_0 so that $H_0 \ge I$; then $R_0 = H_0^{-1}$ is bounded (by 1). It is known that $R_0^{1/2}$ maps $\mathcal{H}$ onto Q_0, and so $R_0^{1/2} X R_0^{1/2}$ is everywhere defined. So, suppose that X is q_0-bounded. Then

$$\begin{aligned} \left| R_0^{1/2} X R_0^{1/2}(\psi, \psi) \right| &= \left| X(R_0^{1/2}\psi, R_0^{1/2}\psi) \right| \\ &\le a q_0(R_0^{1/2}\psi, R_0^{1/2}\psi) + b\|R_0^{1/2}\psi\|^2 \\ &\le (a+b)\|\psi\|^2 \end{aligned}$$

so $R_0^{1/2} X R_0^{1/2}$ is bounded.

For the converse, assume that X is a quadratic form with domain Q_0, and that $a = \|R_0^{1/2} X R_0^{1/2}\| < 1$, and let $\psi = R_0^{1/2}\varphi$ be any element of Q_0. Then

$$\begin{aligned} X(\psi, \psi) &= \langle \varphi, R_0^{1/2} X R_0^{1/2} \varphi \rangle \le a\|\varphi\|^2 \\ &= a q_0(\psi, \psi). \end{aligned}$$

Hence X is q_0-small, with $b = 0$. □

The Kato-Rellich theory can be extended to forms. The key is the *KLMN* theorem, ([35], Vol. 2, p167) which we give here in weaker form.

Theorem 3 *Let H_0 be a positive self-adjoint operator, with quadratic form q_0 and form domain Q_0; let X be a q_0-small symmetric quadratic form. Then there exists a unique self-adjoint operator H_X with form domain Q_0 and such that*

$$\langle H_X^{1/2}\varphi, H_X^{1/2}\psi \rangle = q_0(\varphi, \psi) + X(\varphi, \psi), \qquad \varphi, \psi \in Q_0. \tag{11}$$

Moreover, H_X is bounded below by $-b$, and any domain of essential self-adjointness for H_0 is a form core for H_X.

Now let X be q_0-small, with bounds a, b, with $a < 1-\beta_0$. Denote by H_X the unique operator given by $KLMN$; let q_X denote its form. H_X inherits the main property of H_0, thus:

Lemma 4 $\exp(-\beta H_X)$ *is of trace class for all* $\beta > \beta_X = \beta_0/(1-a)$.

PROOF
We have, as quadratic forms on Q_0, the inequalities

$$-bI + (1-a)q_0 \leq q_X \leq bI + (1+a)q_0. \tag{12}$$

Let L be any finite-dimensional subspace of Q_0, and let q stand for q_0 or q_X. Put

$$\lambda(q, L) = \sup\{q(\psi, \psi) : \|\psi\| = 1, \psi \in L\}. \tag{13}$$

Then the ordered eigenvalues of q are given by the Rayleigh-Ritz principle:

$$\lambda(q, n) = \inf\{\lambda(q, L) : \dim L = n\} \tag{14}$$

From the inequality (12) we have for each subspace L

$$-b + (1-a)\lambda(q_0, L) \leq \lambda(q_X, L) \leq b + (1+a)\lambda(q_0, L). \tag{15}$$

Taking now the inf over L we get [9].

$$-b + (1-a)\lambda(q_0, n) \leq \lambda(q_X, n) \leq b + (1+a)\lambda(q_0, n). \tag{16}$$

Since $\lambda(q_0, n) \to \infty$ as $n \to \infty$, the spectrum of H_X is purely discrete. We thus get for any $\beta > 0$,

$$e^{b\beta}e^{-(1-a)\lambda(q_0,n)\beta} \geq e^{-\lambda(q_X,n)\beta} \geq e^{-b\beta}e^{-(1+a)\lambda(q_0,n)\beta}.$$

Taking the sum over n gives the traces:

$$e^{b\beta}Tr\left(e^{-(1-a)H_0\beta}\right) \geq Tr\left(e^{-H_X\beta}\right) \geq e^{-b\beta}Tr\left(e^{-(1+a)H_0\beta}\right).$$

So if $\exp\{-(1-a)\beta H_0\}$ is of trace-class for all $(1-a)\beta > \beta_0$, then $\exp\{-H_X\beta\}$ is of trace class for all $\beta > \beta_X = \beta_0/(1-a)$. $\square$.

We define the *Cramer class* at ρ_0 to be the q_0-bounded forms X; for then by Lemma (4) there exists a neighbourhood N of zero such that for $\lambda \in N$, $\exp\{-(H_0 + \lambda X\}$ is of trace-class.

It follows from lemma (2) that if $\|R_0^{1/2} X R_0^{1/2}\| < a_o = 1 - \beta_0$, then $H_X > 0$; for we may take $b = 0$ in the *KLMN* theorem.

2.2 The First Piece

We now get the first piece of our manifold. Let $\mathcal{T}(0)$ be the real linear space of q_0-bounded quadratic forms, with domain Q_0 and norm

$$\|X\|_0 = \|R_0^{1/2} X R_0^{1/2}\| < \infty. \tag{17}$$

We note that I, the unit operator, lies in $\mathcal{T}(0)$. The map $A \mapsto H_0^{1/2} A H_0^{1/2}$ from the set of all bounded Hermitian operators $\mathcal{B}(\mathcal{H})$ onto the set of symmetric q_0-bounded sesquiforms is an isometry; this shows that $\mathcal{T}(0)$ is isometric to $\mathcal{B}(\mathcal{H})$; in particular, $\mathcal{T}(0)$ is complete, and so is a Banach space. To each element X of the interior $\mathrm{Int}\,\mathcal{T}_{a_o}(0)$ of the ball in $\mathcal{T}(0)$ of radius $a_o = 1 - \beta_0$, define the density matrix

$$\rho_X = Z^{-1} e^{-(H_0+X)} = e^{-H_0+X+\Psi_X}. \tag{18}$$

The first piece of our manifold is the set $\mathcal{M}_0$ of such states. We set up the patch by mapping $\mathcal{M}_0$ bijectively onto the interior of a ball in a Banach space; our space $\mathcal{T}(0)$, with its ball $\mathcal{T}_{a_o}(0)$, will not do, for if we alter X by adding a multiple of I, we do not change the state; $\rho_X = \rho_{X+\alpha I}$, as the change in X just leads to an equal and opposite change in Z_X, which cancels. Conversely, if $\rho_X = \rho_Y$ lie in $\mathcal{M}_0$, then $X - Y$ is a multiple of I. For, as ρ_X, ρ_Y are faithful states, we may take logarithms: $\log \rho_X = \log \rho_Y$. Then

$$H_0 + X + \Psi_X = H_0 + Y + \Psi_Y$$

so

$$X - Y = (\Psi_Y - \Psi_X) I.$$

Furnish $\mathcal{T}$ with an equivalence relation

$$X \sim Y \qquad\qquad \text{if } X - Y = \alpha I \text{ for some } \alpha \in \mathbf{R}. \tag{19}$$

Then the equivalence classes are lines in $\mathcal{T}(0)$ parallel to I. We furnish the set $\mathcal{T}(0)/\sim$ of equivalence classes with the topology induced from $\mathcal{T}$. That is, an open set in the quotient is the set of all lines passing through some open set in $\mathcal{T}(0)$. There is then a bijection between the set $\mathcal{M}_0$ and the subset of this quotient space defined by

$$\left\{ \{X + \alpha I\}_{\alpha \in \mathbf{R}} : \|X\|_0 < a_o = 1 - \beta_0 \right\}. \tag{20}$$

The bijection is given by

$$\rho_X \mapsto \tilde{\rho}_X = \{Y \in \mathcal{T}(0) : Y = X + \alpha I \text{ for some } \alpha \in \mathbf{R}\}. \tag{21}$$

Thus $\mathcal{M}_0$ becomes topologised, by transfer of structure. It is obviously a Hausdorff space. Indeed, it is well known that the quotient topology is equivalent to the topology given by the quotient norm [21], p140. However, to construct the patch, we parametrise $\mathcal{M}_0$ by a ball in a Banach subspace rather than the quotient. In finite dimensions [14, 27, 31, 33] this has been done by selecting a point on each line in $\widehat{\mathcal{T}}$, namely, the centred variable $\widehat{X} = X - \rho_0.X$. The trouble is that we cannot prove that $\rho_0 X$ is a sesquiform of trace class. We can however find a natural definition for its trace. Suppose that $X \in \mathcal{T}(0)$, and consider the sesquiform $\rho_0^{\lambda} X \rho_0^{\lambda'}$ for $0 < \lambda < 1$. Choose $\beta_1 \in (\beta_0, 1)$ and put $\delta_1 = \lambda(1-\beta_1)$ and $\delta_2 = \lambda'(1-\beta_1)$. From associativity, this is equal to

$$\left(\rho_0^{\lambda-\delta_1}\right)\left(\rho_0^{\delta_1} H_0^{1/2}\right)\left(R_0^{1/2} X R_0^{1/2}\right)\left(H_0^{1/2}\rho_0^{\delta_2}\right)\left(\rho_0^{\lambda'-\delta_2}\right). \qquad (22)$$

This is an operator of trace class, as we see from the Hölder inequality for Schatten norms:

$$\|ABCDE\|_1 \leq \|A\|_{1/\lambda}\|B\|_\infty\|C\|_\infty\|D\|_\infty\|E\|_{1/\lambda'} \qquad (23)$$

where $A \ldots E$ are the factors bracketted in eq. (22). For example, we note that

$$\|A\|_{1/\lambda} = \left(\mathrm{Tr}\left|\rho_0^{\lambda-\delta_1}\right|^{1/\lambda}\right)^{\lambda} = \|\rho_0^{\beta_1}\|_1^{\lambda} < \infty$$

since $\beta_1 > \beta_0$. By cyclicity of the trace, its trace is independent of $\lambda \in (0,1)$. This needs proving, because of the possible existence of Connes cyclic cocycles. Formal differentiation of $\rho_0^{\lambda} X \rho_0^{\lambda'}$ gives

$$\rho_0^{\lambda}[X, H_0]\rho_0^{\lambda'}$$

whose trace might be equal to

$$\mathrm{Tr}\,(\rho_0 X H_0 - H_0 \rho_0 X) = 0$$

by cyclicity; but it is just such expressions that might give non-zero cyclic cocycles if not all the operators are bounded.

We hope to peel off a small part of $\rho_0^{\lambda'}$ and put it at the front. This would be possible if the remaining factor were of trace class. This is proved in the following

Lemma 5 *Suppose that ρ_0^{β} is of trace class for all $\beta > \beta_0$, where $0 \leq \beta_0 < 1$; suppose that X is q_0 bounded. Then*

$$\rho_0^{\lambda} X \rho_0^{\lambda'}$$

is of trace class for all $0<\lambda<1$*, and its trace is independent of* λ*; here as always,* $\lambda'=1-\lambda$*.*

PROOF. The form is that of a bounded operator, since e.g. ρ_0^λ maps $\mathcal{H}$ into Q_0. Write

$$\rho_0^\lambda X\rho_0^{\lambda'}=\rho_0^{\lambda-\delta_1}\left(\rho_0^{\delta_1}H_0^{1/2}\right)\left(R_0^{1/2}XR_0^{1/2}\right)\left(H_0^{1/2}\rho_0^{\delta_2}\right)\rho_0^{\lambda'-\delta_2-\delta}\rho_0^{\delta} \tag{24}$$

for suitably chosen $\delta,\delta_1,\delta_2>0$, such that $\delta_1<\lambda$ and $\delta+\delta_2<\lambda'$. The product of the three operators in brackets in eq. (24) is bounded, by C say (but not uniformly in the δ's). By the Hölder inequality, this gives

$$\left\|\rho_0^\lambda X\rho_0^{\lambda'-\delta}\right\|_1\le\left\|\rho_0^{\lambda-\delta_1}\right\|_{1/(\lambda-\delta_1)}.C.\left\|\rho_0^{\lambda'-\delta_2-\delta}\right\|_{1/\mu}$$

where $\lambda-\delta_1+\mu=1$, i.e. $\mu=\lambda'+\delta_1$. Now, ρ_0 is a positive operator, so

$$\|\rho_0^{\lambda-\delta_1}\|_{1/(\lambda-\delta_1)}=(\mathrm{Tr}\,\rho_0)^{\lambda-\delta_1}=1,$$

and

$$\|\rho_0^{\lambda'-\delta_2-\delta}\|_{1/\mu}$$

is finite if

$$\rho_0^{(\lambda'-\delta_2-\delta)/\mu}$$

is of trace class; the exponent is

$$(\lambda'-\delta_2-\delta)/(\lambda'+\delta_1)$$

which can be made larger than β_0 by choosing the δ's very small. It follows that we can cycle the last factor, ρ_0^δ, to the front without changing the trace, thereby increasing λ and decreasing λ'. We can choose δ as close to $\lambda'-\delta_2$ as we please; it follows that the trace is independent of λ' provided that $\lambda'>\delta_2$. Since δ_2 was arbitrary, we have the result. $\square$

We define the *regularised mean* of X in the state ρ_0 to be

$$\rho_0.X:=\mathrm{Tr}\left(\rho_0^\lambda X\rho_0^{\lambda'}\right),\ \text{for one and hence all } \lambda\in(0,1). \tag{25}$$

Moreover, $\rho_0.X$ is continuous as a map $\mathcal{T}(0)\to\mathbf{R}$. This follows since our bound has $\|X\|_0$ as a factor. The set

$$\widehat{\mathcal{T}}(0):=\{X\in\mathcal{T}(0):\rho_0.X=0\} \tag{26}$$

is a closed linear subspace of $\mathcal{T}$ of codimension 1. The norm $\|\bullet\|_0$, restricted to $\widehat{\mathcal{T}}(0)$, makes $\widehat{\mathcal{T}}(0)$ into a Banach space. We map $\mathcal{M}_0$ bijectively onto the open subset of $\widehat{\mathcal{T}}(0)$ given by the points $\widehat{X}$ where the corresponding points in $\mathcal{T}(0)/\sim$ (i.e. the lines $\tilde{\rho}_X$ parallel to I) intersect it; such a point is unique, being given by $\alpha = -\rho_0.X$. This bijection is a homeomorphism, since both $\mathcal{M}_0$ and $\widehat{\mathcal{T}}(0)$ have been given the topology induced from $\mathcal{T}(0)$. The map $\rho_X \mapsto \widehat{X}$ is our first chart and its inverse is our first patch. As usual in the construction of Banach manifolds, we identify the tangent space at the origin of this chart with the space $\widehat{\mathcal{T}}(0)$ itself. In this identification, the tangent of a curve of the form

$$\{\rho(\lambda) = e^{-(H_0+\lambda X+\Psi_{\lambda X})} : \lambda \in [-\delta, \delta]\}$$

is identified with $\widehat{X} = X - \rho_0.X$. We see from the picture that our patch is the "shadow" of $\mathcal{T}_{1-\beta_0}(0)$ onto the hyperplane $\widehat{\mathcal{T}}(0)$, and that it contains the ball $\widehat{\mathcal{T}}_{1-\beta_0}(0)$ and in its turn is contained in the open set

$$\{Y \in \widehat{\mathcal{T}}_1(0) : \|Y\|_0 < 1 + |\rho_0.Y|\}.$$

We note that in finite dimensions, the set of operators parallel to I is orthogonal to the hyperplane $\widehat{\mathcal{T}}(0)$ when the space $\mathcal{T}(0)$ is furnished with the *BKM* metric. We seem to need more regularity than we have at present if the *BKM* metric is to be finite in infinite dimensions. Obviously, $g_X(Y, I)$ can be defined when one of the operators is the unit operator, and $g_X(Y, I) = \rho_X.Y$. Thus the subspaces $\widehat{\mathcal{T}}(X)$ are all orthogonal to the space parallel to I.

3 Analysis in the First Patch

So far we have a manifold $\mathcal{M}_0$ with one patch. Before enlarging the manifold by the addition of more patches, we do some analysis.

First, it is clear that all states in $\mathcal{M}_0$ have finite entropy and regularised mean energy, which are related by

$$S(\rho_X) = -\mathrm{Tr}\,\rho_X \log \rho_X = \rho_X.H_X + \Psi_X. \tag{27}$$

For $\rho_X.H_X = \mathrm{Tr}\left\{\rho_X^{1-\delta}\left(\rho_X^{\delta} H_X\right)\right\}$ which is finite for $\delta < 1 - \beta_0$.

Lemma 6 *Let A be a closed operator and B be a bounded operator such that $B\mathcal{H} \subseteq D(A)$. Then $C = AB$ is bounded.*

PROOF.
We note that $D(C)=\mathcal{H}$, so by the closed graph theorem it is enough to show that C is closed. For this, let $\psi_n\to\psi$ and suppose that $C\psi_n$ converges. Then we must show that $\psi\in D(C)$ and $C\psi=\lim C\psi_n$. The first is true, as $D(C)=\mathcal{H}$; for the second, we see that $B\psi_n\to B\psi$, as B is bounded, and $A(B\psi_n)$ converges. Since A is closed, we conclude that $B\psi\in D(A)$ (already known) and $C\psi_n=A(B\psi_n)\to AB\psi=C\psi$. □

Lemma 7 *Let $X\in\mathcal{M}_0$, $R_0=H_0^{-1}$ and $R_X=H_X^{-1}$. Then $R_0^{1/2}H_X^{1/2}$ and $R_X^{1/2}H_0^{1/2}$ are bounded.*

PROOF.
Since $H_0\geq I$, we see that $R_0^{1/2}$ is bounded and maps $\mathcal{H}$ into $D(H_0^{1/2})=Q_0=D(H_X^{1/2})$, and $H_X^{1/2}$ is closed. So by lemma (6), $C=H_X^{1/2}R_0^{1/2}$ is bounded; its adjoint, the closure of $R_0^{1/2}H_X^{1/2}$, is therefore also bounded. We find

$$C^*C=R_0^{1/2}H_XR_0^{1/2}=1+R_0^{1/2}XR_0^{1/2},$$

so

$$(1-\|X\|_0)\,I\leq C^*C\leq(1+\|X\|_0)\,I.$$

Thus the inverse of C, namely $H_0^{1/2}R_X^{1/2}$, is bounded by $(1-\|X\|_0)^{-1/2}$. □

Lemma 8 *Let X and Y lie in $\mathcal{M}_0$ and put $q_X=q_0+X$ on Q_0. Then Y is q_X-bounded.*

PROOF.
As sesquiforms, we have

$$\begin{aligned}\left\|R_X^{1/2}YR_X^{1/2}\right\| &= \left\|R_X^{1/2}H_0^{1/2}R_0^{1/2}YR_0^{1/2}H_0^{1/2}R_X^{1/2}\right\|\\ &\leq \left\|R_X^{1/2}H_0^{1/2}\right\|\;\|Y\|_0\;\left\|H_0^{1/2}R_X^{1/2}\right\|\\ &< \infty\end{aligned}$$

by lemma (7). □

We now come to the very useful Duhamel formula for forms.

Theorem 9 *Let X be a symmetric form, q_0-small, and let H_X be the self-adjoint operator with form q_0+X. Then*

$$e^{H_0}-e^{H_X}=\int_0^1 e^{-\lambda H_0}Xe^{-\lambda' H_X}\,d\lambda \tag{28}$$

where the r.h.s. means the limit of $\int_\epsilon^{1-\delta}$ as ϵ and δ converge to zero of the given sesquiform evaluated at any $\psi, \varphi \in \mathcal{H} \times \mathcal{H}$.

PROOF.
Consider the family of operators

$$F(\lambda) = e^{-\lambda H_0} e^{-(1-\lambda) H_X}, \tag{29}$$

where $0 < \lambda < 1$. These are of trace-class, since we can apply Hölder's inequality with parameters $1/\lambda$ and $1/\lambda'$. For any $\psi, \varphi \in \mathcal{H}$ we define

$$F_{\psi,\varphi}(\lambda) = \langle e^{-\lambda H_0} \psi, e^{-(1-\lambda) H_X} \varphi \rangle. \tag{30}$$

Since $e^{-\lambda H_0}$ maps $\mathcal{H}$ into $D(H_0) \subseteq Q_0$, and $e^{-\lambda' H_X}$ maps $\mathcal{H}$ into $D(H_X) \subseteq Q_0$, we see that $F_{\psi,\varphi}$ is differentiable, and

$$\begin{aligned}
\frac{d}{d\lambda} F_{\psi,\varphi}(\lambda) &= -\langle H_0 e^{-\lambda H_0} \psi, e^{-(1-\lambda) H_X} \varphi \rangle + \langle e^{-\lambda H_0} \psi, H_X e^{-(1-\lambda) H_X} \varphi \rangle \\
&= -q_0 \left(e^{-\lambda H_0} \psi, e^{-(1-\lambda) H_X} \varphi \right) + q_X \left(e^{-\lambda H_0} \psi, e^{-(1-\lambda) H_X} \varphi \right) \\
&= X \left(e^{-\lambda H_0} \psi, e^{-(1-\lambda) H_X} \varphi \right).
\end{aligned}$$

Integrating from 0 to 1 gives the theorem. □

Lemma 10 *Suppose that X, Y are q_0-bounded forms, and that the q_0-bound of Y is $a < a_o$. Then $\rho_0^\lambda X \rho_Y^{\lambda'}$ is of trace class for $0 < \lambda < 1$.*

PROOF: We can write

$$\rho_0^\lambda X \rho_Y^{\lambda'} = \rho_0^{\lambda\delta'} \left(\rho_0^{\lambda\delta} H_0^{1/2} \right) \left(R_0^{1/2} X R_0^{1/2} \right) \left(H_0^{1/2} R_Y^{1/2} \right) \left(H_Y^{1/2} \rho_Y^{\lambda'\delta} \right) \rho_Y^{\lambda'\delta'}$$

where $\delta \in (0,1)$ will be chosen soon, and $\delta' = 1 - \delta$. The terms in brackets are all bounded, so the operator norm of their product is bounded by C say; this can grow as λ approaches its limits 0 and 1. We now use Hölder's inequality for traces, to get

$$\begin{aligned}
\left\| \rho_0^\lambda X \rho_Y^{\lambda'} \right\|_1 &\leq C \left\| \rho_0^{\lambda\delta'} \right\|_{1/(\lambda\delta')} \left\| \rho_Y^{\lambda'\delta'} \right\|_{1/\mu} \\
(\text{where } \mu &= \lambda' + \delta\lambda) \\
&\leq C \left\| \rho_0 \right\|_1^{\lambda\delta'} \left\| \rho_Y^{\lambda'\delta'/\mu} \right\|_1^\mu
\end{aligned}$$

which is finite if

$$\left\| \rho_Y^{\lambda'(1-\delta)/(\lambda'+\delta\lambda)} \right\|_1 < \infty.$$

Given λ we can choose δ so small that

$$\frac{\lambda'(1-\delta)}{\lambda'+\delta\lambda} > \beta_Y = \beta_0/(1-a)$$

since the latter is less than 1. $\square$

This does not show that the integral converges in trace norm; for the trace norm of the integrand might become unbounded at the ends, and cannot be shown to be integrable over $[0,1]$. However, the trace, as opposed to the trace-norm, does converge.

Lemma 11 *Let X, Y be q_0-small, Y having bound less than $a_0 = 1-\beta_0$. Then $\mathrm{Tr}\,\rho_0^\lambda X \rho_Y^{\lambda'}$ is bounded for $0<\lambda<1$.*

PROOF.
We first note that we can use the cyclicity of the trace to take out a factor $\rho_0^{\lambda\delta/2}$ on the left of the expression; this is a bounded operator and the remaining product is, as above, still of trace class. We can therefore permute these two factors under the trace. We make use of this when $\lambda \geq 1/2$. If $\lambda < 1/2$ we take out a part of the power of the state ρ_Y from the right and put it on the left under the trace; we now illustrate the method by doing this case.

$$\begin{aligned}\left|\mathrm{Tr}\left(\rho_0^\lambda X \rho_Y^{\lambda'}\right)\right| &= \left|\mathrm{Tr}\left\{\left(\rho_Y^{\lambda'\delta/2} H_Y^{1/2}\right)\left(R_Y^{1/2} H_0^{1/2}\right)\rho_0^\lambda\right.\right.\\ &\qquad \left.\left.\left(R_0^{1/2} X R_0^{1/2}\right)\left(H_0^{1/2} R_Y^{1/2}\right)\left(H_Y^{1/2}\rho_Y^{\lambda'\delta/2}\right)\rho_Y^{\lambda'\delta'}\right\}\right|.\end{aligned}$$

Since $\lambda' \geq 1/2$, we have

$$\left\|\rho_Y^{\lambda'\delta/2} H_Y^{1/2}\right\| \leq \sup_x \left\{x^{1/2} e^{-\delta x/4}\right\} \leq C/\delta^{1/2}$$

This bound occurs twice. The other factors in brackets are bounded operators with norm bounded by C_1, independent of λ and δ. By Hölder,

$$\begin{aligned}\left|\mathrm{Tr}\left(\rho_0^\lambda X \rho_Y^{\lambda'}\right)\right| &\leq C_1 C^2 \delta^{-1}\left\|\rho_0^\lambda\right\|_{1/\lambda}\left\|\rho_Y^{\lambda'\delta'}\right\|_{\lambda'}\\ &\leq C_2\delta^{-1}.1.\left\|\rho_Y^{1-\delta}\right\|_1^{\lambda'}.\end{aligned}$$

Now choose $1-\delta > \beta_Y$ independent of λ. This gives a bound on the trace independent of $\lambda \in (0,1/2)$. Similarly, in the region $\lambda \in (1/2,1)$ we get a bound on the trace independent of λ. $\square$

Corollary 12 *We have*

$$\operatorname{Tr} e^{-H_0} - \operatorname{Tr} e^{-H_X} = \int_0^1 \operatorname{Tr}\left(e^{-\lambda H_0} X e^{-\lambda' H_X}\right) d\lambda. \tag{31}$$

By inserting normalising factors we convert the exponentials into states, and by specialising to the case $X = Y$ we show that the integrand is a bounded function of λ in $(0,1)$. It follows that the integral of the trace is absolutely convergent, and the trace is the sum of the diagonal elements in any orthonormal basis. The trace and the $\int$ can be exchanged, by Fubini's theorem. □

We have obtained an estimate for the perturbation from the state ρ_0; Now H_X inherits the properties of H_0, at least if we replace it by $H_X + I$. We have shown that if Y obeys

$$\|Y\|_X := \|R_X^{1/2} Y R_X^{1/2}\| < a_{\mathrm{x}} \tag{32}$$

then $Y \in \mathcal{M}_0$ is q_X-small; if X is chosen small enough in this norm, and depending on X, Y is also small enough, we may replace H_0 by H_X and H_X by H_{X+Y} in these estimates. This may also be done in lemma (6); this shows

Lemma 13 $\|Y\|_X$ *and* $\|Y\|_0$ *are equivalent norms.*

For,

$$\begin{aligned} \|Y\|_X &= \|R_X^{1/2} H_0^{1/2} R_0^{1/2} Y R_0^{1/2} H_0^{1/2} R_X^{1/2}\| \\ &\leq \|R_X^{12} H_0^{1/2}\|^2 \|Y\|_0 \end{aligned}$$

and the inequality in the other direction is similar. This equivalence is the key to the extension of the manifold to other patches.

We see from the resulting estimate

$$Z_X - Z_{X+Y} = \operatorname{Tr} e^{-H_X} - \operatorname{Tr} e^{-H_{X+Y}} \leq C\|Y\|_X \tag{33}$$

that the partition function Z_X is a Lipschitz function of $X \in \mathcal{M}_0$.

4 Affine Geometry in $\mathcal{M}_0$

By an *affine structure* for a manifold $\mathcal{M}_0$ we mean a rule for forming the convex mixture "$\lambda\rho_1 + \lambda'\rho_2$" $(0 \leq \lambda \leq 1; \rho_1, \rho_2 \in \mathcal{M}_0)$. An affine space is

a space with a specified affine structure; it is necessarily convex. The unit ball $\widehat{\mathcal{T}}_1(0)$ in $\widehat{\mathcal{T}}(0)$ is a convex subset of a Banach space and so has a natural affine structure coming from the linear structure. By 'transfer of structure', the chart $X \mapsto \rho_X$ from $\widehat{\mathcal{T}}_1(0)$ to $\mathcal{M}_0$ provides $\mathcal{M}_0$ with the induced affine structure. This is called the canonical or (+1)-affine structure. Clearly, "$\lambda\rho_X + \lambda'\rho_Y$" $= \rho_{\lambda X+\lambda' Y}$ which differs from the usual mixture of states $\rho = \lambda\rho_X + \lambda'\rho_Y$. The latter is called the mixture or (-1)-affine structure of the state space. While it is obvious that $\mathcal{M}_0$ is (+1)-convex, it is not clear that it is (-1)-convex. That is, while the (-1)-mixture ρ of ρ_1 and ρ_2 is certainly a state, it might not lie in $\mathcal{M}_0$ even if $\rho_1, \rho_2 \in \mathcal{M}_0$.

4.1 The (+1)-affine connection

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be affine spaces; then an affine map $U : \mathcal{T}_1 \to \mathcal{T}_2$ is a map obeying

$$U\left(\lambda\rho_1 + (1-\lambda)\rho_2\right) = \lambda U\rho_1 + (1-\lambda)U\rho_2 \tag{34}$$

for all $\rho_1, \rho_2 \in \mathcal{T}_1$ and $0 \le \lambda \le 1$. An affine connection on a Banach manifold is an assignment, for each continuous curve L from ρ_1 to ρ_2, of an affine map U_L from the tangent space at ρ_1 to the tangent space at ρ_2, obeying $U_L U_{L'} = U_{L\cup L'}$; the map U for the empty path $\emptyset$, when $\rho_1 = \rho_2$, is the identity, and the symbol $L \cup L'$ denotes the path L followed by the path L'; if L' is the path L with reversed parameter, we take $L \cup L' = \emptyset$. These axioms ensure that U_L is an invertible map for any L. An affine connection is linear if $U_L(0) = 0$ for every curve L; a linear connection is torsion-free. If U_L is independent of L then the connection is called *flat*, or curvature-free. The commonly used formulation is the infinitesimal version of the above, obtained by differentiating, if the structure is smooth.

In order to define the (+1)-affine connection concretely, we first put coordinates on the tangent space at any $X \in \mathcal{M}_0$. We have seen that $\rho_0.Y$ is continuous in $Y \in \mathcal{M}_0$. Since H_X inherits all relevant properties of H_0, we obtain a similar estimate $|\rho_X.Y| \le \text{const}\|Y\|_X$. The set

$$\widehat{\mathcal{T}}(X) := \{Y : \rho_X.Y = 0\} \tag{35}$$

is therefore a closed subspace of $\mathcal{T}(0)$ in the equivalent topology defined by $\|\bullet\|_X$. We identify the tangent space at ρ_X to be the Banach space $\widehat{\mathcal{T}}(X)$ with the norm $\|Y\|_X$. We then take the (+1)-parallel transport U_L of $Y - \rho_0.Y \in \widehat{\mathcal{T}}(0)$ along any path L in the manifold to be the point $Y - \rho_X.Y \in \widehat{\mathcal{T}}(X)$. This map takes $Y = 0$ to zero, and extends to a

linear mapping from $\widehat{\mathcal{T}}(0)$ onto $\widehat{\mathcal{T}}(X)$. We see that parallel transport is nothing other than the moving of the representative point in the line $\tilde{\rho}$ from one hyperplane to the other. Since this transport is independent of L and linear, the connection is flat and torsion-free.

5 Extension of the Manifold

We see two ways of extending the manifold by gluing new patches. The first is to try to include as many q_0-small perturbations X as possible, and not just those obeying $\|X\|_0 < a_o$; recall that this condition is sufficient for X to have q_0-bound less than $a_o < 1$. The second, and main, method of extension, is to use any point ρ_X in the first patch, and to consider perturbations Y of H_X with $\|Y\|_X < a_{\mathrm{x}} = 1 - \beta_X$. This might lead out of $\mathcal{M}_0$; we can continue indefinitely, starting at H_Y etc. In this way we include eventually a state of any temperature, and the manifold points generally in the $+H_0$ direction.

Suppose then that X is symmetric and q_0-small enough. Then there exists a self-adjoint operator H_0 whose form is $q_0 + X$ with form-domain Q_0 and

$$|X(\psi,\psi)| \leq a\left(q_0(\psi,\psi) + b\|\psi\|^2/a\right) \text{ for some } a < a_o. \tag{36}$$

Let $\tilde{H}_0 = H_0 + \frac{b}{a}I$; this is self-adjoint on $D(H_0)$ and

$$|X(\psi,\psi)| \leq a\langle {\tilde{H}_0}^{1/2}\psi, {\tilde{H}_0}^{1/2}\psi\rangle.$$

Let $\tilde{R}_0 = {\tilde{H}_0}^{-1}$. Then for $\psi \in Q_0$, we have for $\psi \in \mathcal{H}$,

$$\tilde{R}_0^{1/2} X \tilde{R}_0^{1/2}(\psi,\psi) \leq a\langle \tilde{R}^{1/2}\psi, \tilde{H}_0\tilde{R}_0^{1/2}\psi\rangle = a\|\psi\|^2, \tag{37}$$

so

$$\|X\|_{\tilde{0}} := \left\|\tilde{R}_0^{1/2} X \tilde{R}_0^{1/2}\right\| \leq a < a_o.$$

We have thus shown that for any q_0-small form X with bound $< a_o$ there exists a choice of Hamiltonian $\tilde{H}_0$ such that ρ_X lies inside the open ball $\|X\|_{\tilde{0}} < a_o$. Let us furnish $\widehat{\mathcal{T}}(0)$ with this norm; it is equivalent to the norm $\|X\|_0$, since $\tilde{R}_0^{1/2}H_0^{1/2}$ and $R_0^{1/2}\tilde{H}_0^{1/2}$ are both bounded. We can therefore add the patches got in this way to the first patch, to get a Banach manifold. We can enlarge the manifold even further, by analogy with the classical case [33]. Let X be a q_0-small form; it therefore defines a unique pair of self-adjoint operators, $H_\pm$, with forms $q_\pm := q_0 \pm X$; we include ρ_X in the first patch if for each choice of sign, $\rho_\pm = \exp -H_\pm$ is of trace class. For such an X the quantum analogue of the Luxemburg norm is finite.

Definition 14 *We put*

$$\|X\|_L = \inf\left\{r > 0 : \mathrm{Tr}\left[\left(\exp-(H_0 + X/r) + \exp-(H_0 - X/r)\right)/(2Z_0)\right] - 1 < 1\right\}.$$

For large r, X/r is q_0-small, and so the traces make sense; since Z_X is continuous in X/r if it is small, the set in the infimum is non empty. As r becomes smaller, either the operator $H_\pm$ fails to be unique, or the finiteness of the trace might fail; in either case we put the trace equal to ∞, and the corresponding r is a lower bound for $\|X\|_L$. It can be shown that $\|X\|_L$ is a seminorm.

The second extension of the manifold is to construct a similar chart around a state ρ_X as we did around ρ_0, where X is $\tilde{q}_0$-small, with bound $< a_o$. Since the choice of H_0 was anyway arbitrary provided $H_0 \geq I$, we drop the tilde; so we consider $X \in \mathcal{M}_0$. Choose $H_X \geq I$. This Hamiltonian inherits all the properties of H_0. Let Y be q_X-small with bound $< a_{\mathrm{x}}$; then there is a unique self-adjoint operator H_{X+Y}, whose form domain is Q_0, and whose form is $q_0 + X + Y$, such that

$$\rho_{X+Y} = Z^{-1}_{X+Y} e^{-H_{X+Y}} \tag{38}$$

is of trace-class, and $\rho_X.Y$ can be defined as $\mathrm{Tr}\left(\rho_X^\lambda Y \rho_X^{\lambda'}\right)$. Let $\mathcal{T}(X)$ be the Banach space of forms Y such that $\|Y\|_X < \infty$, with this norm. Since $\|\bullet\|_0$ and $\|\bullet\|_X$ are equivalent norms, this space is actually the same as $\mathcal{T}(0)$ as a set. The interior of this ball in $\mathcal{T}_X$ consists of certain q_X-small forms which are q_0-bounded but might not lie in $\mathcal{M}(0)$. Let $\mathcal{M}_X$ be the set of states of the form eq. (38). Again, two forms that differ by a multiple of I yield the same state, and there is a bijection between $\mathcal{M}_X$ and the set of lines $\{\tilde{\rho}_{X+Y}\}$ in $\mathcal{T}_X$ parallel to I that cut the open ball in $\mathcal{T}(X)$ of radius $1 - \beta_X$. The set of such lines is furnished with the quotient topology. Let $\widehat{\mathcal{T}}(X)$ be the (closed) hyperplane $\{T \in \mathcal{T}(X) : \rho_X.Y = 0\}$. Each line $\tilde{\rho}_{X+Y}$ cuts this hyperplane in a unique point, and those in the neighbourhood of $Y = 0$ cut the plane inside our open ball $\widehat{\mathcal{T}}_{a_{\mathrm{x}}}(X)$. This gives a chart from an open set in $\mathcal{M}_X$ onto this ball. Again, the tangent space at ρ_X is identified with $\widehat{\mathcal{T}}(X)$. The $(+1)$-affine structure in $\mathcal{M}_X$ is that induced from the linear structure of $\mathcal{T}(X)$. We can enlarge this piece of the manifold to include all q_X-small perturbations Y such that $\exp-\{H_0 + X + Y\}$ is of trace class, and can cover the enlarged set of states by consistent overlapping charts, in the same way as for the first method of extension of $\mathcal{M}_0$.

The next step in building the manifold is to consider the union of $\mathcal{M}_0$ and $\mathcal{M}_X$. The two charts are topologically compatible, in that in the overlap

$\mathcal{M}_0 \cup \mathcal{M}_X$ the two norms $\| \bullet \|_0$ and $\| \bullet \|_X$ induced by the charts are equivalent; see lemma (13). The (+1)-affine structure of $\mathcal{M}_0$ and $\mathcal{M}_X$ are the same on their overlap, since both are induced by the linear structure of $\mathcal{T}(0)$. Our choice of parallel transport with the first patch reflects this, and can be extended in stages to a transport between any two points in the union of the pieces. Indeed, let $\widehat{X}_1$, $\widehat{X}_2$ lie in $\widehat{\mathcal{T}}(0)$, so their means in ρ_0 are zero. Put $\widehat{Z} = \lambda\widehat{X}_1 + \lambda'\widehat{X}_2$, and let U denote the parallel transport from ρ_0 to ρ_X. Then

$$U\widehat{X}_i = \widehat{X}_i - \rho_X.\widehat{X}_i, \;\; i = 1,2, \text{ and } U\widehat{Z} = \widehat{Z} - \rho_X.Z$$

Then

$$\begin{aligned} U\left(\lambda\widehat{X}_1 + \lambda'\widehat{X}_2\right) &= \widehat{Z} - \rho_X.\widehat{Z} \\ &= \lambda\widehat{X}_1 + \lambda'\widehat{X}_2 - \rho_X.\left(\lambda\widehat{X}_1 + \lambda'\widehat{X}_2\right) \\ &= \lambda U\widehat{X}_1 + \lambda' U\widehat{X}_2. \end{aligned}$$

That is, U takes the convex mixture in $\mathcal{T}(0)$ to that in $\mathcal{T}(X)$. Thus the union of the first two pieces is a Banach manifold furnished with a flat torsion-free affine structure and the (+1)-parallel transport U.

We can extend further, to a third piece, starting from a different point X' in $\mathcal{M}_0$ or from a point in $\mathcal{M}_X$ outside $\mathcal{M}_0$. In either case we arrive at a q_0-bounded form with domain Q_0, and a third piece of the manifold with a chart into an open ball of the Banach space $\{Y : \rho_{X'}.Y = 0\}$, with norm $\| \bullet \|_{X'}$ equivalent to the norms already defined. We continue by induction, starting at any point in the manifold obtained already, to get to any q_0-bounded form that can be arrived at in a finite number of steps. At each stage, starting from ρ_X we enlarge the ball of radius a_{x} by the first method, to include all q_X-small forms which define a state. Moreover, suppose we arrive at two far points, $H_0 + X$ and $H_0 + Y$, which however lie in each other's patch. When neither X nor Y is q_0-small (but are q_0-bounded), we can by construction find a finite chain $X_1, X_2 \ldots$ from X to H_0 and another finite chain $Y_1, Y_2 \ldots$ from H_0 to Y, each small relative to the last; then $R_X^{1/2} Y R_X^{1/2}$ is a finite product

$$R_X^{1/2} H_{X_1}^{1/2}\, R_{X_1}^{1/2} H_{X_1}^{1/2} \ldots H_0^{1/2} R_{Y_1}^{1/2}\, H_{Y_1}^{1/2} R_{Y_2}^{1/2} \ldots R_{Y_n}^{1/2} H_Y R_{Y_n}^{1/2} \ldots R_X^{1/2}$$

which is bounded. Thus $\|Y\|_X$ is finite. Similarly $\|X\|_Y$ is finite. Thus when we then extend to all states obtainable in this way in a finite number

of steps, all the norms of any overlapping region are equivalent. With each enlargement, we define the (+1)-affine structure and parallel transport in stages from chart to chart, to give a flat torsion-free connection.

Definition 15 *The information manifold $\mathcal{M}$ defined by H_0 consists of all states obtainable in a finite number of steps, by extending from $\mathcal{M}_0$ by either the first method or the second method, as explained above.*

The Cramer class of each $\rho \in \mathcal{M}$ is the set of q_0-bounded forms.

The question now arises, when we add perturbations $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ as above, and $X_1 + \ldots + X_n = Y_1 + \ldots + Y_m$ as forms (on Q_0), whether we arrive at the same state whichever route we take. We do, since there is a unique self-adjoint operator defined by the form

$$q_0 + X_1 + \ldots + X_n = q_0 + Y_1 + \ldots + Y_m$$

with form domain Q_0.

We now have a natural result.

Theorem 16 *$\mathcal{M}$ is (+1)-convex.*

PROOF. We first prove the result when $\beta_0 = 0$. Then the only condition on the size of a perturbation Y of H_X is that it be q_X-small. In this case it is obvious that the manifold is a cone.

Let $\mathcal{M}(0)$ denote the set of states ρ_X where X is q_0-small. Then we define $\mathcal{M}(n)$ inductively by

Definition 17 *$\rho_X \in \mathcal{M}(n)$ if there exists $\rho_Y \in \mathcal{M}(n-1)$ such that $X - Y$ is q_Y-small, where $q_Y = q_0 + Y_1 + \ldots + Y_{n-1}$, and each addition Y_j is small relative to q_{j-1}.*

We show that $\mathcal{M}(0)$ is (+1)-convex, and that if $\mathcal{M}(n-1)$ is (+1)-convex, so is $\mathcal{M}$.

Suppose then that $X_i \in \mathcal{M}_0$, $1 = 1, 2$. Then for $\psi \in Q_0$,

$$\begin{aligned} |X_1(\psi,\psi)| &\leq a_1 q_0(\psi,\psi) + b_1\|\psi\|^2 \\ |X_2(\psi,\psi)| &\leq a_2 q_0(\psi,\psi) + b_2\|\psi\|^2. \end{aligned}$$

Then

$$\begin{aligned} \left|(\lambda X_1 + \lambda' X_2)(\psi,\psi)\right| &\leq \lambda\,|X_1(\psi,\psi)| + \lambda'\,|X_2(\psi,\psi)| \\ &\leq \lambda\left(a_1 q_0(\psi,\psi) + b_1\|\psi\|^2\right) + \lambda'\left(a_2 q_0(\psi,\psi) + b_2\|\psi\|^2\right) \\ &\leq a q_0(\psi,\psi) + b\|\psi\|^2 \end{aligned}$$

where $a = \max\{a_1, a_2\} < 1$ and $b = \max\{b_1, b_2\}$. Hence $\lambda X_1 + \lambda' X_2$ is q_0-small, and $\mathcal{M}(0)$ is convex.

Now let $\mathcal{M}(n)$ be obtained from $\mathcal{M}(n-1)$ as in the definition (15). So let q_Y be of the form $q_0 + Y$ where $\rho_Y \in \mathcal{M}(n-1)$, and let X be q_Y-small. Then $\rho_{X+Y} \in \mathcal{M}(n)$, and any element of $\mathcal{M}(n)$ is of this form. Let $\rho_1, \rho_2 \in \mathcal{M}(n)$; then there exist Y_1, Y_2 such that $\rho_{Y_1}, \rho_{Y_2} \in \mathcal{M}(n-1)$, and writing $q_1 = q_0 + Y_1$ and $q_2 = q_0 + Y_2$, then there exist forms X_1, X_2 such that X_1 is q_1-small and X_2 is q_2-small, and ρ_1, ρ_2 are the states corresponding to $q_1 + X_1$ and $q_2 + X_2$. Let $q = \lambda q_1 + \lambda' q_2$; the state corresponding to q lies in $\mathcal{M}(n-1)$, since this is (+1)-convex, by the inductive hypothesis. A simple estimate shows that $\lambda X_1 + \lambda' X_2$ is q-small, so that the state corresponding to $q + \lambda X_1 + \lambda' X_2$ lies in $\mathcal{M}(n)$. But the latter is $\lambda(q_1 + X_1) + \lambda'(q_2 + X_2)$, whose corresponding state is the (+1) mixture of ρ_1 and ρ_2. This shows that $\mathcal{M}(n)$ is (+1)-convex.

Now relax the condition that $\beta_0 = 0$, and define the part $\mathcal{M}(n)$ to be the set of states obtained from $\mathcal{M}(n-1)$ as above, but using only sufficiently small perturbations. Then all the conclusions derived above remain true, up to the result that $\lambda X_1 + \lambda' X_2$ is q-small. Thus $q + \lambda X_1 + \lambda' X_2$ is the form of a self-adjoint operator that is bounded below; call this operator H. Now, by the convexity of Z_X, $\exp -H$ is of trace class, since ρ_1 and ρ_2 are. The same is true if we replace X_i by $-X_i$; hence $Z^{-1} \exp -X$ lies in $\mathcal{M}(n)$, by the first method of extension. $\square$

We have not been able to prove that the manifold is (−1)-convex; if ρ_1 and ρ_2 are density operators in the first patch, then obviously, $\rho := \lambda\rho_1 + \lambda'\rho_2$ is a density operator. All we can show from the operator convexity of $-\log x$ [5, 24], is that $-\log \rho = H_0 + X$, where X has q_0-bound 1; but we need the bound to be *less than* 1 for ρ to lie inside the first patch.

6 Outlook

We have defined a Banach manifold, with the flat torsion-free (+1)-connection. The canonical variables at ρ_0, are the centred q_0-bounded forms X, with the norm $\|\bullet\|_0$. These are (+1)-affine coordinates, and the manifold is a convex set when expressed in terms of these. The Massieu function Ψ is a continuous convex function on the manifold. We can therefore construct the Legendre transform using Fenchel's theory, to obtain the 'mixture' variables $\rho_Y.X$ at any point ρ_Y in the manifold. The entropy is a continuous function. With more regularity, we have been able to show that the *BKM* metric is finite at

regular points, and is the Fréchet derivative of $\rho_Y.X$, as in the classical and finite-dimensional cases. Moreover, the free energy is real-analytic. This work, [13, 40] which extends [37], will be published elsewhere.

Acknowledgements

This work is part of an ongoing programme being worked out with P. Combe, G. Burdet and H. Nencka, of CPT and CNRS Marseille and the University of Madeira. I thank P. Combe for arranging the visit to Marseille, and CNRS for financial support; I thank H. Nencka for arranging the visit to Madeira. I am endebted to E. B. Davies for help with the proof of Lemma 4.

References

[1] S. Amari, **Differential Geometric Methods in Statistics**, Lecture Notes in Statistics, **28**, Springer, Berlin, 1985.

[2] S. Amari, *Information Geometry*, in **Geometry and Nature**, Ed. J.-P. Bourguignon and H. Nencka, Contemporary Physics **203**, 1997.

[3] H. Araki, *Relative Hamiltonians for Faithful Normal States of a von Neumann Algebra*, Publ. R. I. M. S. (Kyoto), **9**, 165-209, 1973. *ibid*, **A4**, 361-, 1968.

[4] R. Balian, Y. Alhassid and H. Reinhardt, *Dissipation in many-body systems: a geometrical approach based on information theory*, Physics Reports, **131**, 2-146, 1986.

[5] J. Bendat and S. Sherman, Trans. Amer. Math. Soc., **79**, 58-71, 1955.

[6] N. N. Bogoliubov, Phys. Abh. Sov. Union, **1**, 229-, 1962.

[7] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **72**, 3439-3443, 1994.

[8] N. N. Chentsov, *Statistical Decision Rules and Optimal Inferences*, in Russian, 1972; Translations of Mathematical Monographs, **53**, Amer. Math. Soc., Providence, 1982.

[9] E. B. Davies, **Spectral Theory of Differential Operators**, Camb. Univ. Press, 1995, Th. 4.5.1 and Th. 4.5.3.

[10] A. P. Dawid, *Discussion of a paper by B. Effron*, Annals of Statistics, **3**, 1231-1234, 1975. *ibid*, *Further comments on a paper by Bradley Effron*, **5**, 1249-, 1977.

[11] R. A. Fisher, *Theory of statistical estimation*, Proc. Camb. Phil. Soc., **22**, 700-725, 1925.

[12] P. Gibilisco and G. Pistone, *Connections on Nonparametric Statistical Manifolds by Orlicz Space Geometry*, Infin. Dim. Analysis, Quantum Probability **1**, 325-347, 1998, World Scientific. — and T. Isola, *Connections on statistical manifolds of density operators by geometry of noncommutative L^p spaces*, *ibid*, **2**, 169-178, 1999.

[13] M. R. Grasselli and R. F. Streater, *The quantum information manifold for epsilon bounded forms*, math-ph/9910031.

[14] H. Hasagawa, *α-Divergence of the non-commutative Information Geometry*, Reports on Math. Phys., **33**, 87-93, 1993.

[15] H. Hasagawa, *Noncommutative extension of information geometry*, pp 327-337 in **Quantum Communication and Measurement**, Eds. V. P. Belavkin, O. Hirota and R. L. Hudson, Plenum Press, 1995.

[16] H. Hasagawa, *Exponential and Mixture Families in Quantum Statistics: Dual Structure and Unbiased Parameter Estimation*, Reports on Math. Phys., **39**, 49-68, 1997.

[17] H. Hasagawa and D. Petz, *Non-commutative Extension of Information Geometry*, pp 109-118 in **Quantum Communication, Computing and Measurement**, Ed. O. Hirota et al, Plenum Press, N.Y. 1997.

[18] R. S. Ingarden, *Information theory and variational principles in statistical physics*, Bull. acad. polon sci. Math-Astro-Phys, **11**, 541-547, 1963. —-, *Information thermodynamics and differential geometry*, Memoirs Sagami Inst. Techn., **12**, 83-89, 1978. —-, and H. Janyszek, *On the local Riemannian structure of the state space of classical information thermodynamics*, Tensor, **39**, 279-285, 1982. —-, *Towards mesoscopic thermodynamics*, Open Systems and Info. Dynamics, **1**, 75-102, 1992; —-, *Information geometry in functional spaces; classical and quantum finite statistical systems*, Int. J. Engineering Sci., **19**, 1609-1633, 1981. R. S. Ingarden, Y. Sato, K. Sagura and T. Kawaguchi, *Information thermodynamics and differential geometry*, Tensor, **33**, 347-353, 1979;

R. S. Ingarden, H. Janyszek, A. Kossakowski and T. Kawaguchi, *Information geometry of quantum statistical systems*, Tensor, **37**, 105-111, 1982.

[19] R. S. Ingarden and T. Nakagomi, *The second order extension of the Gibbs state*, Open Systems and Info. Dyn., **1**, 243-258, 1992.

[20] E. T. Jaynes, *Information Theory and Statistical Mechanics* Phys. Rev., **106**, 620-630, and II, *ibid* **108**, 171-190, 1957.

[21] T. Kato, **Perturbation Theory for Linear Operators**, Springer-Verlag, 1966.

[22] A. Kossakowski, *On the Quantum Informational Thermodynamics* Bull. acad. polonaise des sci., **17**, 263-267, 1969.

[23] R. Kubo, J. Phys. Soc. Japan, **12**, 570, 1957. Also, *The fluctuation-dissipation theorem*, Reports on Prog. in Phys., **29**, 255-284, 1966.

[24] G. Lindblad, *Entropy, information and quantum measurements*, Comm. Math. Phys., **33**, 305-322 (1973).

[25] T. Matsubara, Prog. Theor. Phys.,**14**, 351, 1955.

[26] H. Mori, Prog. Theor. Phys., **33**, 423, 1965.

[27] H. Nagaoka, *Differential geometrical aspects of quantum state estimation and relative entropy*, pp 449-452 in **Quantum Communication and Measurement**, Eds. V. P. Belavkin, O. Hirota and R. L. Hudson, Plenum Press, 1995.

[28] M. Ohya and D. Petz, **Quantum Entropy and its Use**, Springer, Heidelberg, 1993.

[29] D. Petz, *Geometry of canonical correlation on the state space of a quantum system*, J. Math. Phys., **35**, 780-795, 1994.

[30] D. Petz, *Monotone Metrics on Matrix Spaces*, Linear Algebra and its Applications, **244**, 81-96, 1996.

[31] D. Petz and G. Toth, *The Bogoliubov Inner Product in Quantum Statistics*, Lett. Math. Phys., **27**, 205-216, 1993.

[32] D. Petz, and H. Hasagawa, *On the Riemannian Metric of α-Entropies of Density Matrices*, Lett. in Math. Phys., **38**, 221-225, 1996.

[33] G. Pistone and C. Sempi, *An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one*, The Annals of Statistics, **33**, 1543-1561, 1995.

[34] C. R. Rao, *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math Soc., **37**, 81-89, 1945.

[35] M. Reed and B. Simon, **Methods of Modern Mathematical Physics**, Academic Press, Vol. 2, 1975.

[36] G. Roepstorff, *Correlation Inequalities in Quantum Statistical Mechanics and their Application in the Kondo Problem*, Comm. Math. Phys., **46**, 253-262, 1976. See also **Path-Integral Approach to Quantum Physics**, Springer, Berlin, 1994.

[37] R. F. Streater, *Information Geometry and Reduced Quantum Description*, Reports on Mathematical Physics, **38**, 419-436, 1996.

[38] R. F. Streater, **Statistical Dynamics**, Imperial College Press, 1995.

[39] R. F. Streater, *Statistical Dynamics and Information Geometry*, pp 117-131 in **Geometry and Nature**, Eds. J.-P. Bourguignon and H. Nencka, Contemporary Mathematics **203**, AMS, 1997.

[40] R. F. Streater, *The analytic quantum information manifold*, to appear in **Stochastic Processes, Physics and Geometry: new interplays**, eds. F. Gesztesy, S. Paycha and H. Holden, Can Math. Soc. math-ph/9910036

[41] A. Uhlmann, Reports on Math. Phys., **9**, 273, 1976; *Density operators as an arena for differential geometry*, *ibid*, **33**, 253-263, 1993.